Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.081458

ARTICLE

Toward Secure and Adaptive Medical Digital Twins: A Privacy-Preserving Federated Multi-Agent Reinforcement Learning Framework

Tallha Akram[1,*], **Sadiq Ahmad**[2,*] **and Meshal Alharbi**[3]

[1]Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

[2]COMSATS University Islamabad, Wah Campus, Electrical Engineering Department, Wah Cantt, Pakistan

[3]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

*Corresponding Authors: Tallha Akram. Email: t.akram@psau.edu.sa; Sadiq Ahmad. Email: engrsadiqahmad@ciitwah.edu.pk

Received: 02 March 2026; Accepted: 12 June 2026; Published: 13 August 2026

ABSTRACT: Scalability limitations, privacy risks, and lack of adaptability remain key challenges in centralized medical digital win (MDT) architectures. While federated learning (FL) mitigates the need to share raw data, it often lacks adaptability to dynamic clinical environments and does not fully integrate formal privacy guarantees into the learning process. To address these challenges, this paper proposes a decentralized, federated, multi-agent reinforcement learning (F-MARL) framework to coordinate MDTs in the presence of partial observability. The framework is formulated as a multi-agent partially observable Markov decision process (MA-POMDP), enabling distributed policy optimization in heterogeneous and uncertain clinical settings. We introduce a novel algorithm, privacy-aware advantage actor–Critic with personalization and privacy protection (PA3C-PP), which integrates (i) differentially private gradient perturbation, (ii) weighted federated aggregation, and (iii) adaptive global–local policy fusion for personalization. Unlike many existing healthcare-oriented federated learning approaches that treat privacy mainly as an external or post-optimization mechanism, the proposed method incorporates differential privacy within the reinforcement learning update process. Experimental evaluation in a smart-ICU simulation demonstrates improved learning stability, enhanced resilience to agent failures, and stronger privacy protection, while maintaining latency and operational costs comparable to non-private federated baselines. These findings indicate that privacy-aware federated reinforcement learning provides a promising direction for scalable, adaptive, and regulation-compliant decentralized healthcare intelligence.

KEYWORDS: Medical digital twin; federated multi-agent reinforcement learning; PA3C-PP; MA-POMDP

1 Introduction

Intensive care units (ICUs) continuously generate streams of physiological and clinical data, such as heart rate, oxygen saturation, ventilator parameters, infusion rates, and other patient-specific measurements [1]. Medical digital twins (MDTs) are virtual patients that enable real-time monitoring, prediction, and decision support by continuously synchronizing clinical data with computational models [2,3]. Deploying these models in vivo in real-world settings requires generalizable learning systems that are adaptive, scalable, and privacy-preserving.

Recent developments in artificial intelligence (AI), IoT, and cloud-edge computing facilitate intelligent healthcare systems for monitoring, predictive diagnosis, and personalized treatment [4]. Many MDT architectures still rely on centralized data processing and cloud-based coordination, leading to latency,

communication overload, scalability issues, a single point of failure, and the risk of privacy leakage [5]. In healthcare, there are many regulations on patient privacy, such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR) [6].

Federated learning (FL) alleviates the need to directly share raw data across institutions or devices by training a shared model across participants. However, common toolkits designed for conventional supervised learning may not be sufficiently adaptable for real-time clinical decision-making. Multi-agent reinforcement learning (MARL) is a natural paradigm in which previous studies have considered decentralized sequential decision-making, with each agent learning a policy through interaction with its local environment. However, applying MARL-based methods to healthcare MDTs raises concerns about privacy leakage from shared updates, communication overhead, heterogeneous clinical conditions, and difficulty achieving convergence.

To overcome the above-mentioned hurdles, we devise a privacy-preserving federated multi-agent reinforcement learning (F-MARL) for MDT coordination. We conceptualize the MDT as an autonomous learning agent operating under partial observability, with global coordination achieved through privacy-protected federated aggregation. The proposed privacy-aware advantage actor-critic with personalization and privacy protection (PA3C-PP) algorithm integrates differentially private gradient perturbation, weighted federated aggregation, and adaptive global-local policy fusion. Instead of treating privacy as a post-processing step externally tacked on after the actor-critic training, PA3C-PP takes direct ownership of the privacy protection scheme.

Fig. 1 shows a simple comparison between centralized MDT coordination, where patients' data/ecosystem and decision-making itself are all aggregated to a cloud optimizer, increasing risks and latencies in terms of privacy exposure and possible failure, vs. an application of F-MARL, where MDT agents learn locally, distributing knowledge and sharing only privacy-protected model updates.

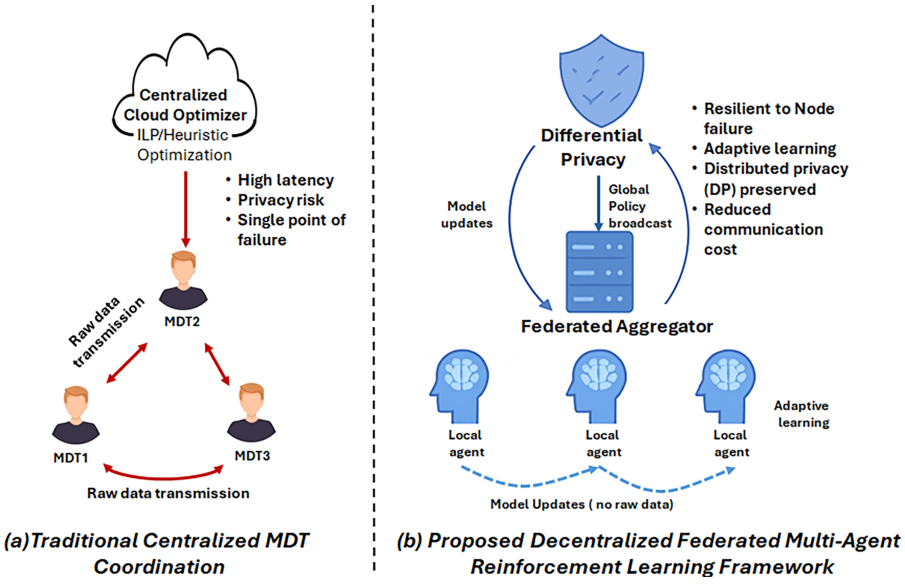

Figure 1: Conceptual overview of Medical Digital Twin (MDT) coordination paradigms. (**a**) Traditional centralized MDT coordination. (**b**) Proposed decentralized Federated Multi-Agent Reinforcement Learning (F-MARL) framework.

The key contributions of this work can be summarized as follows:

1. A PA3C-PP algorithm for MDT coordination, injecting differential privacy in actor-critic RL updates.
2. A personalized federated aggregation approach that achieves a meaningful trade-off of global coordination and local adaptation under heterogeneous and non-IID clinical settings.
3. A stability-preserving federated multi-agent learning method constructed through the combination of advantage-based policy optimization, privacy-aware updates, and weighted aggregation.
4. The method is validated in a smart-ICU simulation environment, demonstrating enhanced learning stability, privacy-aware performance, and robustness relative to selected baselines.

For better readability, we define the main mathematical symbols the first time we use them while listing the full nomenclature in Appendix A.

The structure of the remainder of this article is set out as follows. A review of related research is presented in Section 2. Section 3 presents the proposed framework and the PA3C-PP algorithm. Simulation setup and results are described in Section 4. Finally, a summary of the work contained in this article is provided in Section 5.

2 State-of-the-Art Solutions

This section briefly surveys related work in MDT systems, healthcare FL, and RL/MARL for distributed medical decision-making, highlighting the key technical gaps that motivate the proposed PA3C-PP algorithm, particularly in decentralization, adaptability, personalization, and privacy-aware learning.

2.1 Medical Digital Twin Systems

Digital twin technologies show promise in personalized and predictive healthcare. For examples of MDT use for real-time monitoring, predictive diagnosis, and virtual patients, as reviewed by the authors in [2] and in [3]. Many MDT methods are cloud-based, with clinical data being stored or processed on remote servers. These approaches are simple to manage but can suffer from latency, scalability issues, a single point of failure, and exposure of private data. These issues have led to decentralized methods of coordinating MDTs that allow for local learning and reduce the transfer of sensitive patient data.

2.2 Federated Learning in Healthcare

FL allows common model training without sharing raw data between individual institutions in the healthcare ecosystem. It has shown benefits for privacy-preserving health care analysis and for learning from electronic health records [7–9]. However, most healthcare FL systems still rely on centralized aggregation, leading to communication delays, a single point of failure (SPoF), and possibly information leakage in the exchanged model updates [10–12]. Some hybrid DT-FL systems have been studied for distributed hospital analytics and edge-cloud workload sharing [13,14], although most methods target flat prediction tasks rather than adapting sequential decision-making tasks.

Recent personalized FL techniques demonstrate that performing model averaging across both global and local models can yield significantly greater robustness to non-IID data distributions [15,16]. Similar to personalization, differential privacy has been an established tool used in FL pipelines to reduce information leakage from shared updates [17–19]. The latter focuses on how to apply privacy protection in FL, mainly as an external mechanism or as post-processing of updates, rather than as part of the learning update, thereby motivating a privacy-aware design for federated reinforcement learning.

2.3 Reinforcement and Multi-Agent Learning in Healthcare

RL can be applied to clinical problems involving sequential decision-making under uncertainty, such as treatment scheduling, resource allocation, imaging use, and ICU case management. MARL extends this ability by enabling multiple agents to learn coordinated policies in distributed settings. Recent federated MARL methods include heterogeneous-aware or meta-learning-based approaches for collaborative learning, such as FedMRL and FedMetaMed [20]. However, these still suffer from limitations in privacy calibration, fault tolerance, personalization, and stable learning in partially observable and communicable settings.

As summarized in Table 1, existing work provides important building blocks for MDTs, FL, and MARL-based healthcare intelligence. However, it often considers decentralization, privacy preservation, personalization, and adaptive policy learning separately. In contrast, our proposed PA3C-PP framework integrates these aspects into a unified F-MARL formulation, in which MDT agents learn adaptive policies locally, exchange privacy-protected updates, and personalize global knowledge under heterogeneous clinical conditions.

Table 1: State-of-the-art solutions in digital twin, federated learning, and reinforcement learning for healthcare systems.

Ref.	Domain/Technique	Key Contribution	Limitation/Gap
[2]	Digital twin review	Reviews DT applications in healthcare	Limited focus on decentralized coordination
[3]	Scoping review	Summarizes DT use cases and readiness levels	Does not address distributed learning
[7]	Healthcare FL	Reviews FL for healthcare informatics	Mainly centralized aggregation
[13]	Federated DT	Combines DT and FL for hospital analytics	Limited adaptive policy optimization
[8]	Privacy-aware FL	Incorporates DP for EHR-based learning	Lacks agent-level autonomy
[14]	DT-based control	Studies edge-cloud collaboration for DT control	Centralized structure and limited adaptability

3 Proposed Framework

In this section, we introduce the proposed F-MARL framework for decentralized MDT coordination and privacy-preserving decision-making. The framework leverages local actor-critic reinforcement learning, weighted federated aggregation, personalized global-local model fusion, and differential privacy. The MDT coordination problem is formulated as a multi-agent partially observable Markov decision process (MA-POMDP) in order to support distributed decision-making under partial observability and heterogeneous clinical conditions.

3.1 System Overview

As shown in Fig. 2, each patient is represented by an MDT connected to IoT-enabled physiological and clinical data streams. Each MDT hosts a local RL agent that learns patient-specific decision policies using local observations. Instead of transmitting plain clinical data, MDTs send privacy-protected actor-critic model updates to a federated aggregator, which performs weighted model averaging and returns a

global policy that is then fused with each local policy for local personalization. Privacy protection based on differential privacy is applied on-site to prevent leakage of patient information from shared updates. Secure aggregation and communication, anomaly detection, and robust aggregation against attacks such as model poisoning, membership inference, and reconstruction should also be included in practical deployments.

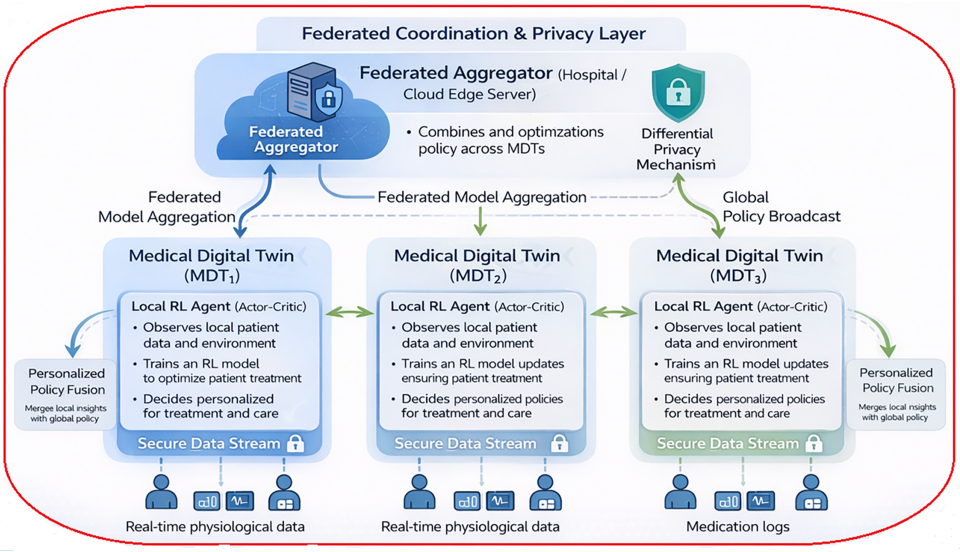

Figure 2: Overview of the proposed Federated Multi-Agent Reinforcement Learning (F-MARL) framework.

In this sense, differential privacy and federated learning offer important privacy-preserving techniques, but regulatory compliance in healthcare systems involves more than merely computing privacy guarantees. Our proposed framework partially supports data minimization by only maintaining raw physiological and clinical data at a healthcare institution and transmitting privacy-protected model updates. Our task-specific federated learning workflow would also partially align with the purpose limitation, since model updates would be generated only for specified clinical coordination objectives. However, a real HIPAA/GDPR-compliant deployment would still require complementary governance systems for obtaining patient consent, managing institutional access-control policies, maintaining audit logs, implementing secure data-retention procedures, and building compliance-monitoring infrastructure. As such, differential privacy should be considered as part of the healthcare governance framework rather than a complete solution.

3.2 Mathematical Formulation

The decentralized MDT coordination task is modeled as an MA-POMDP:

$$\mathcal{M} = \langle \mathcal{N}, \mathcal{S}, \{\mathcal{A}_i\}, \{\mathcal{O}_i\}, P, R, \gamma \rangle, \tag{1}$$

where $\mathcal{N}$ denotes the set of MDT agents, $\mathcal{S}$ is the global state space, $\mathcal{A}_i$ and $\mathcal{O}_i$ are the action and observation spaces of agent i, P is the transition function, R is the reward function, and γ is the discount factor. Each agent follows a stochastic policy $\pi_i(a_t|o_t;\theta_i)$ parameterized by actor parameters θ_i and supported by critic parameters v_i. The local objective is

$$J_i(\theta_i) = \mathbb{E}_{\pi_i}\left[\sum_{t=0}^{T} \gamma^t R_i(s_t, a_t)\right], \tag{2}$$

and the actor-critic update uses the advantage estimate.

$$\nabla_{\theta_i} J_i = \mathbb{E}_t\left[\nabla_{\theta_i} \log \pi_i(a_t|o_t;\theta_i) A_i(t)\right], \qquad A_i(t) = R_i(t) + \gamma V_i(s_{t+1}) - V_i(s_t). \tag{3}$$

Here, $A_i(t)$ reduces policy-gradient variance and supports stable learning under partially observable MDT environments.

3.3 Federated Aggregation and Personalized Model Fusion

At the communication round r, each MDT performs local actor-critic learning and sends privacy-protected model updates to the aggregator. The global actor and critic parameters are obtained using weighted federated averaging:

$$\theta^{(r+1)} = \sum_{i=1}^{N} \frac{n_i}{\sum_j n_j}\left(\theta_i^{(r)} + \mathcal{N}(0, \sigma^2 I)\right), \qquad v^{(r+1)} = \sum_{i=1}^{N} \frac{n_i}{\sum_j n_j}\left(v_i^{(r)} + \mathcal{N}(0, \sigma^2 I)\right), \tag{4}$$

where $\theta_i^{(r)}$ and $v_i^{(r)}$ are the local actor and critic parameters, n_i is the local participation weight, and $\mathcal{N}(0, \sigma^2 I)$ denotes Gaussian perturbation for differential privacy. The perturbation is applied after gradient clipping and before transmission, ensuring that only privacy-protected updates are exchanged.

After aggregation, each MDT personalizes the received global model by fusing it with its local model:

$$\theta_i^{\text{fused}} = (1-\eta)\theta_i^{\text{local}} + \eta\theta^{(r+1)}, \qquad v_i^{\text{fused}} = (1-\eta)v_i^{\text{local}} + \eta v^{(r+1)}, \tag{5}$$

where $\eta \in [0,1]$ controls the balance between local specialization and global coordination. A larger value of η gives more weight to the global model, whereas a smaller value of η preserves stronger local adaptation. This interpolation-based personalization follows the general principle of personalized FL, where local models are regularized toward shared global representations while retaining client-specific knowledge. Similar global-local fusion strategies have been adopted in personalized FL methods such as FedPer [21], APFL [22], and mixture-based personalized federated optimization [23].

The fusion mechanism can also be interpreted through the following regularized objective:

$$\min_{\theta_i} (1-\eta) f_i(\theta_i) + \eta f_{\text{global}}(\theta), \tag{6}$$

where $f_i(\cdot)$ denotes the local objective and $f_{\text{global}}(\cdot)$ represents the global federated objective. This formulation supports heterogeneous and non-IID MDT environments by balancing local empirical risk with global knowledge consistency.

3.4 PA3C-PP Algorithm and Differential Privacy Calibration

The PA3C-PP algorithm shown in Algorithm 1 combines local advantage actor-critic learning, personalized federated aggregation, and differential privacy. Each MDT agent computes local learned policy updates, clips the update norm to bound sensitivity, adds a Gaussian perturbation at the vector level, and sends only the privacy-protected update to the aggregator. The initial value of the clipping threshold C is computed as the median of the gradient norms observed during early training rounds and is then updated within a bounded interval to remove deleterious outliers while preserving useful gradients.

Each local update g_i is clipped as

$$\bar{g}_i = g_i \cdot \min\left(1, \frac{C}{\|g_i\|_2}\right), \tag{7}$$

and the privacy-protected update is computed as

$$\tilde{g}_i = \bar{g}_i + \mathcal{N}(0, \sigma^2 C^2 I). \tag{8}$$

Vector-level perturbation is used instead of independent scalar-level perturbation to reduce dimensionality-related noise amplification and stabilize training. Following standard DP-SGD practice [24], the Gaussian noise scale satisfies

$$\sigma \geq \frac{C\sqrt{2\ln(1.25/\delta)}}{\varepsilon}, \tag{9}$$

where (ε, δ) are the differential privacy parameters. The privacy-utility trade-off is governed jointly by C and σ: smaller values of C and larger values of σ improve privacy but may reduce learning utility.

To estimate cumulative privacy loss over R communication rounds, the proposed framework adopts an advanced-composition-based approximation:

$$\varepsilon_{\text{tot}} \leq \sqrt{2R\ln(1/\delta)}\,\frac{C}{\sigma} + R\frac{C^2}{\sigma^2}. \tag{10}$$

This bound follows the basic advanced-composition intuition of treating each clipped and Gaussian-perturbed federated update as one privacy-relevant mechanism so that we approximate the cumulative privacy loss as a square-root composition term plus a linear higher-order accumulation term over R communication rounds.

Algorithm 1: PA3C-PP for privacy-preserving federated MDT coordination

1: Initialize global actor $\theta^{(0)}$ and critic $v^{(0)}$
2: **for** each communication round $r = 0, 1, \ldots, R-1$ **do**
3: **for** each participating MDT agent $i \in \mathcal{N}$ **do**
4: Receive global parameters $\theta^{(r)}, v^{(r)}$
5: Fuse global and local parameters using Eq. (5)
6: Perform local actor-critic updates using local observations
7: Clip local gradients using Eq. (7)
8: Add Gaussian noise using Eq. (8)
9: Send privacy-protected updates to the federated aggregator
10: **end for**
11: Aggregate actor and critic parameters using Eq. (4)
12: Update cumulative privacy budget using Eq. (10)
13: **end for**
14: **return** personalized MDT policies $\{\theta_i^{\text{fused}}, v_i^{\text{fused}}\}_{i=1}^{N}$

3.5 Stability and Complexity Analysis

There are three mechanisms within PAC3-PP designed to improve the stability of the learning process: 1) Advantage-based gradient estimation reduces the variance of the updates. 2) Federated synchronization

reduces the chances of policy drift amongst MDT agents, and 3) Personalized fusion of global and local estimates regularizes the learning process when working under heterogeneous clinical environments. Differential privacy noise may introduce stochasticity, but vector-level perturbation and norm clipping help preserve stable aggregate updates.

For m participating agents, K local training steps, and actor-critic parameter dimensions d_θ and d_v, the local computational cost per agent is

$$T_{\text{local}} = \mathcal{O}\left(Kc_{\text{fb}}(d_\theta + d_v)\right), \tag{11}$$

where c_{fb} is the cost of one forward/backward pass. The communication cost per round is

$$T_{\text{comm}} = \mathcal{O}\left(m(d_\theta + d_v)b\right) \tag{12}$$

where b denotes the bit precision used to transmit model parameters. Thus, local computation scales with model size and the number of local training steps, while communication overhead scales with the number of participating agents and the parameter dimensionality. For larger hospital-network deployments, hierarchical aggregation, asynchronous updates, partial client participation, and parameter compression can reduce synchronization overhead.

4 Simulation and Discussion

This section evaluates the proposed PA3C-PP framework using a smart-ICU simulation environment. The simulation models decentralized MDT coordination among patient-monitoring agents, heterogeneous compute nodes, and privacy-preserving federated learning rounds. The objective is to evaluate learning performance, latency, operational cost, privacy protection, convergence stability, and robustness under controlled clinical-computing conditions.

4.1 Smart-ICU Simulation Setup

The simulator consists of $N = 5$ MDT agents connected to heterogeneous edge, fog, and cloud compute nodes. Each MDT represents a patient-monitoring unit that generates physiological data streams and computational tasks, such as real-time analysis, anomaly detection, and treatment-support recommendations. Synthetic physiological streams were generated using statistical characteristics derived from open-source ICU datasets such as PhysioNet [25]. This design provides a reproducible and privacy-safe environment for evaluating decentralized coordination; however, it does not fully capture all higher-order dependencies, rare pathological events, or institutional biases present in real-world ICU trajectories.

Each MDT agent selects a compute node for task execution under dynamic workload, latency, and resource conditions. The reinforcement learning components are defined as follows:

- **State (s_t):** patient status indicator, task queue length, compute-node availability, and network latency.
- **Action (a_t):** task allocation to one of the available compute layers:

$$a_t \in \{\text{Edge}, \text{Fog}, \text{Cloud}\}. \tag{13}$$

- **Reward (J):** a tri-objective utility balancing cost, latency, and security:

$$J = -\left(\alpha \cdot \text{Cost} + \beta \cdot \text{Latency} - \lambda_s \cdot \text{Security}\right). \tag{14}$$

In Eq. (14), α, β, and λ_s control the relative importance of operational cost, latency, and security, respectively. In the default smart-ICU configuration, $(\alpha, \beta, \lambda_s) = (1, 5, 3)$ is used. The latency term receives the largest weight because timely response is critical in acute-care environments, while λ_s emphasizes privacy protection and regulatory compliance. The symbol λ_s denotes the security weight and is distinct from the reinforcement learning discount factor γ.

All experiments were conducted for $R = 100$ federated communication rounds with $K = 10{,}000$ local update steps per round. The main hyperparameters were actor learning rate $\alpha_a = 10^{-4}$, critic learning rate $\alpha_c = 10^{-3}$, discount factor $\gamma = 0.99$, personalization coefficient $\eta = 0.8$, clipping threshold $C = 1.0$, and Gaussian noise scale $\sigma = 0.5$. All compared methods used the same actor-critic architecture and training schedule, while only the aggregation, personalization, and privacy mechanisms were varied. Results were averaged over five independent simulation runs with different random seeds.

4.2 Benchmarking Methods

The proposed F-MARL/PA3C-PP framework is compared with representative baselines to distinguish the effects of decentralized coordination, personalization, and privacy preservation:

- **NF-MARL (Local Only):** Each MDT agent learns independently without federated aggregation. This baseline shows the effect of removing collaborative learning.
- **FedAvg-MARL:** a standard federated MARL approach using classical FedAvg without global-local personalization.
- **DP-FedAvg:** a differentially private federated averaging baseline where Gaussian noise is added to model updates following DP-SGD principles, without reinforcement-learning-specific personalization.
- **F-MARL (Non-Private):** a federated MARL baseline without differential privacy. This provides a non-private upper-performance reference for evaluating the privacy-utility trade-off.
- **Centralized MARL:** a centralized training setting where learning information is globally available. This represents an idealized non-private upper-bound scenario.

These baselines provide a broader comparison beyond ablation-style variants and help clarify the individual contributions of personalization, decentralized coordination, and differential privacy.

4.3 Evaluation Metrics

To comprehensively evaluate performance, four complementary metrics are considered:

1. **Average tri-objective reward (J):** normalized utility combining latency, operational cost, and security according to Eq. (14). Higher reward indicates better overall policy quality.
2. **Average latency (ms):** mean task completion time across MDT agents; lower values indicate faster responsiveness.
3. **Average operational cost ($):** average monetary proxy associated with task offloading and node selection; lower values indicate more economical coordination.
4. **Average security score (S):** normalized privacy score derived from the cumulative differential privacy loss:

$$S = \exp(-\epsilon_{\text{tot}}), \tag{15}$$

where ϵ_{tot} denotes the total accumulated privacy budget over all communication rounds. This definition ensures that a lower privacy loss corresponds to a higher security score (i.e., greater privacy preservation), with $S \in (0, 1]$, a score closer to 1 indicating better security. We average our reported results across independent simulation runs with different seeds and report the mean performance and

standard deviation over the last 20 communication rounds, which capture the replication and stability of convergence.

As indicated in Table 2, the F-MARL/PA3C-PP method achieves the highest security score and the least oscillation, while keeping latency and cost in the range of the non-private F-MARL. The centralized MARL baseline shows slightly lower latency/costs as it does not preserve distributed privacy or federated constraints. In contrast, the PA3C-PP enforces strong privacy with marginal operational overhead. The method also attains a slightly higher average reward than the non-private federated baseline; this could be attributed to the regularizing effect of the personalized global-local fusion, and both induce stochasticity, resulting in lower policy variance in the heterogeneous MDT.

Table 2: Performance and convergence stability comparison of evaluated methods (mean ± standard deviation over the last 20 communication rounds).

Method	Reward (J)	Latency (ms)	Cost ($)	Security	Oscillation
NF-MARL (Local Only)	−6.21 ± 0.42	58.7 ± 2.8	3.24 ± 0.11	0.41 ± 0.05	0.125
FedAvg-MARL	−3.35 ± 0.21	45.6 ± 2.1	2.89 ± 0.10	0.52 ± 0.04	0.098
DP-FedAvg	−3.55 ± 0.24	47.8 ± 2.4	2.93 ± 0.12	0.72 ± 0.03	0.121
F-MARL (Non-Private)	−3.08 ± 0.17	42.3 ± 1.9	2.75 ± 0.09	0.56 ± 0.03	0.083
Centralized MARL	−2.85 ± 0.15	40.8 ± 1.7	2.70 ± 0.08	0.48 ± 0.02	0.083
F-MARL/PA3C-PP (Proposed)	**-2.94 ± 0.13**	**43.0 ± 1.5**	**2.76 ± 0.06**	**0.78 ± 0.02**	**0.047**

Fig. 3 compares convergence stability with the standard deviation of the tri-objective reward across the last 20 communication rounds. Lower values represent smoother, more stable convergence behavior. Our proposed F-MARL/PA3C-PP framework exhibits the lowest oscillation among the compared methods, indicating more stable convergence under privacy-preserving federated coordination. NF-MARL and DP-FedAvg produce larger oscillations due to less-stable local policy updates, making convergence more vulnerable to stochastic variation.

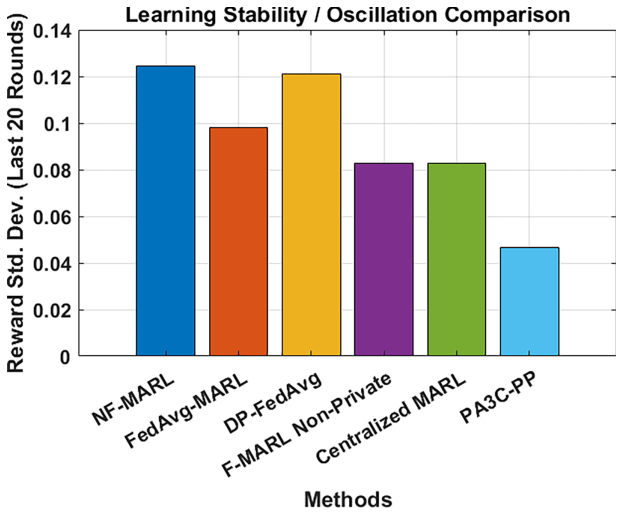

Figure 3: Oscillation comparison across evaluated baseline methods. Lower values indicate more stable convergence behavior.

4.4 Decentralized Efficiency and Adaptivity

Fig. 4 tracks average tri-objective reward vs. rounds of communication. The reported reward values are dimensionless normalized utility scores of the weighted sum of latency, cost, and security objectives. The proposed PA3C-PP method converges more smoothly than NF-MARL and DP-FedAvg, showing the stabilizing influence of federated coordination and personalized global-local policy mixing.

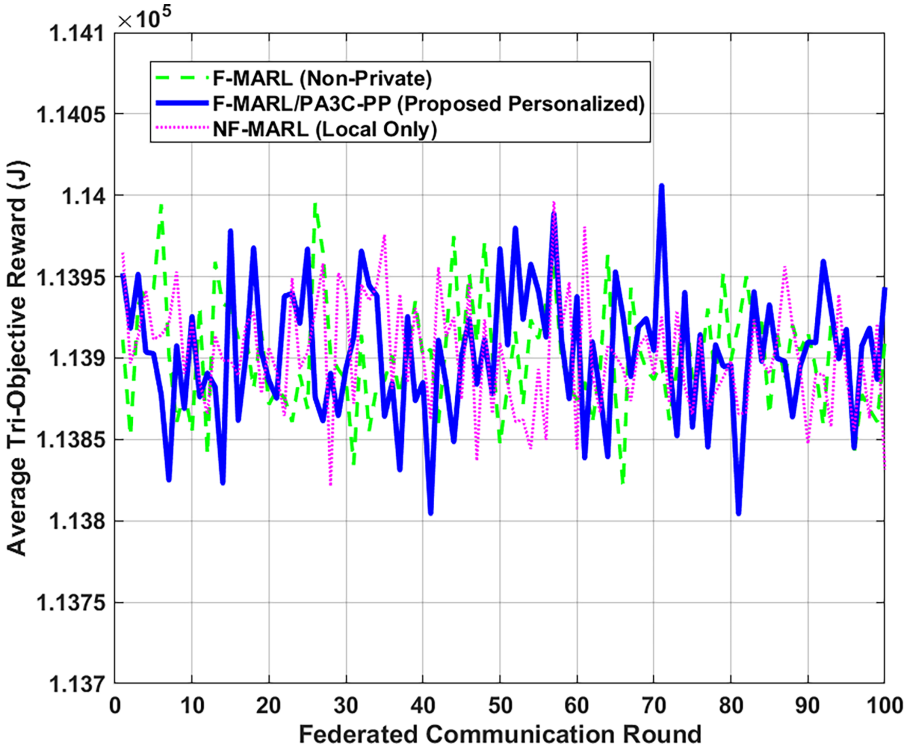

Figure 4: Average tri-objective reward convergence across federated communication rounds. Curves represent mean performance over multiple independent simulation runs (dimensionless normalized tri-objective reward).

We define convergence as the number of communication rounds it takes to reach 95% of the maximum observed reward. Under this criterion, F-MARL/PA3C-PP takes about 20% fewer communication rounds than the selected baselines, averaged over independent runs. Thus, convergence speed is evaluated quantitatively rather than merely by looking at the reward curves.

4.5 Privacy-Utility Trade-Off

Fig. 5 illustrates the trade-off between latency and operational cost for the different coordination methods. The non-private F-MARL baseline is slightly lower on both axes because it has no privacy noise to account for. PA3C-PP remains close to the dot but yields a substantially better security score, indicating that vector-wise differential-privacy perturbation and personalized aggregation are sufficient to preserve learning utility while improving privacy protection.

The privacy-utility trade-off is governed by the mixing factor (clipping threshold C) and the Gaussian noise scale σ. Smaller C reduces gradient sensitivity and strengthens privacy but may cut off useful signals. Larger σ improves privacy protection but increases stochasticity and thus lowers convergence speed. In PA3C-PP, personalized fusion somewhat mitigates privacy-induced noise; however, it also allows the MDT to keep policy information locally useful.

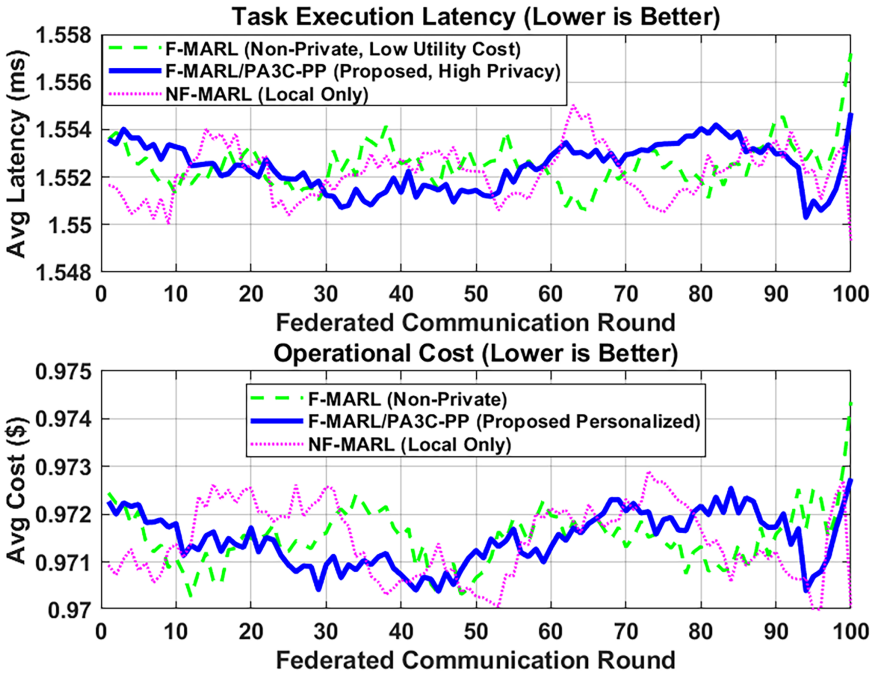

Figure 5: Privacy-utility trade-off. Top: average task execution latency. Bottom: average operational cost across federated communication rounds.

4.6 Resilience and Security Analysis

Fig. 6 assesses the behavior of the framework in response to a "controlled" node-failure experiment (Round 50). The temporary surging cost suggests that they adapt to the loss of a single participating node, with stability indicating that the weighted aggregation and local policy adjustment compensate for the simulated failure. The security comparison further suggests that PA3C-PP facilitates a higher privacy-aware security score than the non-private and local-only baselines. The experiment simulates a simplified single-node abrupt-outage scenario, whereas real hospital networks exhibit correlated failures, intermittent connectivity, partial communication loss, gradual degradation, and more. As such, this experiment should be seen as merely an initial controlled resilience test and not a complete fault-tolerance validation.

4.7 Hyperparameter Sensitivity Analysis

The performance of our framework depends on the personalization coefficient η, the clipping threshold C, and the noise scale σ, which together govern our privacy, utility, and convergence behavior.

Effect of η: This parameter $\eta \in [0,1]$ trades off global model aggregation and local user personalization, with higher values of aggregation tending to give better convergence stability (due to adding more global shared knowledge and coordination), but lower values tending to give better adaptation to heterogeneous or totally non-IID clinical conditions.

Effect of C: The clipping threshold C, which globally bounds the norm of the local gradients, limits sensitivity for differential privacy. A smaller value of C provides stronger privacy guarantees but may distort gradient information too much, whereas a larger value of C preserves more utility but requires stronger noise to achieve the same privacy.

Effect of σ: The noise scale σ dictates the magnitude of the Gaussian noise added for privacy preservation. Increasing σ increases privacy guarantees, but unless small noise perturbations are sufficiently amplified, the learning process may become too stochastic and unstable.

Interaction effects: The interactions between η, C, and σ are also important; for example, higher noise levels are likely to require stronger global coordination to achieve stability, whilst lower clipping thresholds require careful tuning of the noise scale to avoid excessive loss of utility. Our current experiments, with $\eta \in [0.6, 0.9]$, yielded a good, balanced trade-off across the evaluated scenarios, but the balance may differ across emergency care, chronic care, and privacy-sensitive institutional deployments.

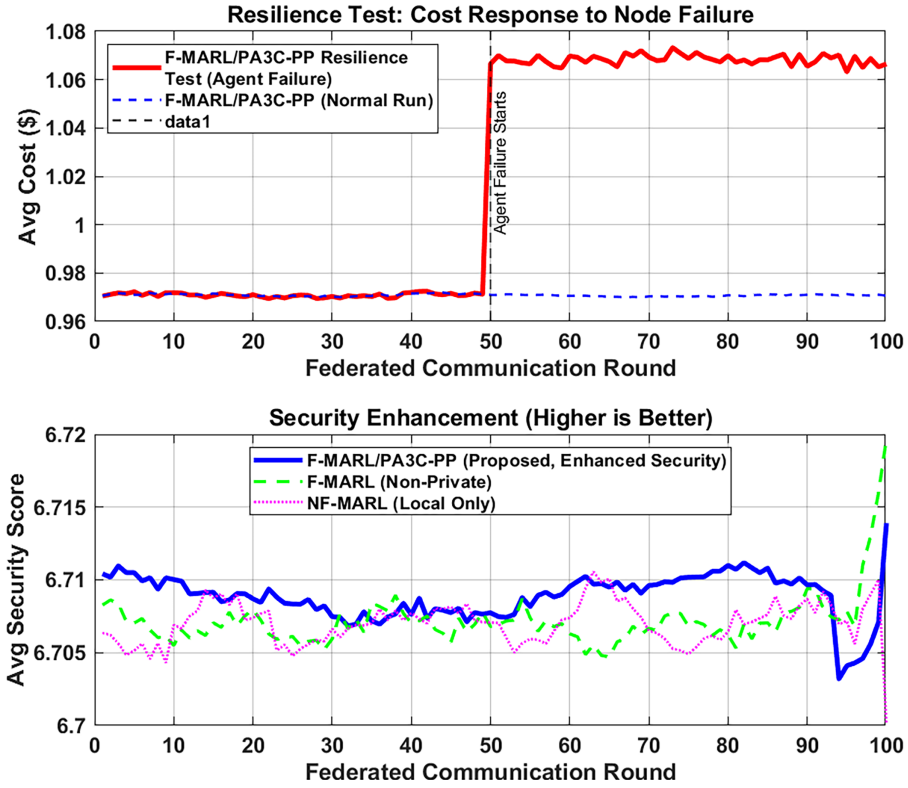

Figure 6: Resilience and security analysis. Top: cost response to a node failure at Round 50. Bottom: comparative security score across communication rounds.

4.8 Robustness under Heterogeneity and Non-IID Conditions

In our PA3C-PP framework, we focus on robustness in heterogeneous MDT within practical healthcare systems through personalized federated aggregation, where each MDT agent observes different patient populations, disease distributions, monitoring frequencies, workloads, communication conditions, and so on. Such heterogeneity leads to policy divergence and unstable convergence for the standard federated learning approaches.

The personalization function mitigates these effects by allowing each MDT agent to retain locally specialized policy behavior while still drawing on knowledge from a global model. In our current smart-ICU simulation, non-IID behavior is approximated by workload, resource availability, communication delays, and synthetic physiological characteristics shared among MDT agents. Hence, our results indicate robustness under simulated moderate heterogeneity rather than universal non-IID clinical distributions.

4.9 Scalability and Communication Analysis

From a scalability perspective, local actor-critic learning can be performed independently at each MDT, and only model parameters (not raw clinical data) are exchanged via the aggregator, alleviating the direct burden of data transmission and enabling privacy-preserving collaboration. A 100-round training session with five MDT agents and 10,000 local steps per round took roughly 32 min in the current settings, suggesting small-scale feasibility.

The communication overhead per federated round can be expressed as

$$T_{\text{comm}} = \mathcal{O}\left(m(d_\theta + d_v)b\right), \tag{16}$$

where m is the number of participating agents, d_θ and d_v are the actor and critic parameter dimensions and **b** denotes the transmission bit precision. Overall, communication cost increases with model dimensionality, number of clients, and frequency of aggregation. In larger hospital networks, we may need to resort to methods such as hierarchical aggregation, asynchronous updates, partial client participation, adaptive aggregation intervals, parameter compression, and gradient sparsification.

4.10 Comprehensive Discussion

Overall, the simulation results demonstrate that PA3C-PP enhances decentralized MDT coordination through a synergistic integration of adaptive reinforcement learning, privacy-aware federated aggregation, and personalized global-local fusion. Relative to the local-only and non-personalized baselines, the proposed coordination achieves stronger privacy protection, smoother convergence, and competitive latency and cost. Our results further show that privacy preservation does not necessarily entail significant performance losses when personalization and stable aggregation are incorporated into the learning process.

This work can be considered a proof-of-concept experiment/validation. Our simulation adopts synthetic PhysioNet-derived data, 20 MDT agents, and a too-simple 3-layer edge-fog-cloud compute model. While the node-failure and convergence-stability results appear robust, hospital network disruptions might be further complicated by a maximum-bandwidth constraint, missing data, adversarial behavior, and non-static patient-condition perturbations. We hope to further validate our framework via real ICU datasets $\in$ MIMIC-IV; larger MDT populations ($N > 20$); broader partitioning via non-IID partitions; dynamic resource models; multimodal patient data; and stronger threat models (poisoning, reconstruction, and mem-inference attacks).

5 Conclusion

This study proposed a decentralized, privacy-preserving coordination framework for medical digital twins (MDTs) based on federated multi-agent reinforcement learning (F-MARL). The proposed PA3C-PP algorithm integrates advanced actor-critic learning, personalized federated aggregation, and differential privacy within a unified learning framework. By allowing each MDT agent to learn locally while exchanging only privacy-protected model updates, the framework supports adaptive decision-making without direct sharing of sensitive physiological or clinical data.

The smart-ICU simulation results show that PA3C-PP improves decentralized MDT coordination in four main aspects: (i) faster and smoother convergence through advantage-based learning and personalized global-local fusion; (ii) stronger privacy protection through gradient clipping and Gaussian perturbation; (iii) competitive latency and operational cost compared with non-private federated baselines; and

(iv) improved resilience under the considered node-failure scenario. These results suggest that privacy-aware federated reinforcement learning is a promising direction for scalable and regulation-conscious healthcare intelligence.

The personalization coefficient η played a key role in achieving model consistency vs. local adaptation, particularly under heterogeneous, non-IID clinical conditions. Likewise, the differential privacy mechanism provided actor-critic-level protection while enabling satisfactory convergence. While the present results should be treated as proof-of-concept validation rather than evidence for clinical deployment, we note that our evaluation is based on synthetic PhysioNet-derived data and limited MDT actors ($N = 5$) and thus may not fully exhibit rare clinical events, higher-order temporal dependencies or rhythms, institutional bias, or all potential forms of non-IID patient distributions.

There are a number of practical limitations. The simulator uses a simple edge-fog-cloud architecture with fixed node capacities, whereas hospital networks may have dynamic resource availability, fluctuating workloads, variable bandwidth, intermittent connectivity, and dynamic resource allocation. The node-failure experiment we conducted is for a controlled, single, isolated, sudden failure, whereas real failures may be correlated, partially connected, isolated, or completely degraded. While differential privacy limits information leakage, it is not a panacea and cannot defend against model poisoning, Byzantine clients, membership inference, or reconstruction attacks.

Future work will validate PA3C-PP using additional large-scale real-world ICU datasets, such as MIMIC-IV. We will also investigate validating PA3C-PP across larger populations of MDTs (e.g., populations for which $N > 20$) and other heterogeneous non-IID partitions; use of dynamic resource models, stochastic failure processes, asynchronous communication, hierarchical aggregation, parameter compression, and multimodal patient data (e.g., imaging, laboratory records, genomics) will also be included in future work. Investigating more intensive approaches to systematic hyperparameter and runtime optimization for η, C, and σ will also be ongoing, as will work on more secure mechanisms such as secure aggregation, stronger aggregation rules, anomaly detection, encoded parameter exchange, and potentially adversarially robust federated optimization.

Acknowledgement: The authors extend their appreciation to the Prince Sattam bin Abdulaziz University for supporting this research work. The authors used an AI-assisted language editing tool to improve the manuscript's clarity, grammar, and readability; all scientific content, analysis, and conclusions remain the authors' sole responsibility.

Funding Statement: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/35090).

Author Contributions: Tallha Akram and Sadiq Ahmad, created the paper's fundamental idea. Tallha Akram and Sadiq Ahmad, have developed the methodology. Meshal Alharbi and Sadiq Ahmad have analyzed the paper. Original draft was prepared by Tallha Akram which was edited by Sadiq Ahmad and Meshal Alharbi. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data, simulator configuration files, random seed settings, hyperparameter configurations, and scripts supporting the findings of this study are available from the corresponding authors upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1: Nomenclature.

Symbol	Meaning
Sets and Indices	
$\mathcal{N}$	Set of MDT agents (clients) participating in the federated system; index $i \in \mathcal{N}$
$\mathcal{S}$	Global state space of the MA-POMDP environment
$\mathcal{A}_i$	Action space of agent i
$\mathcal{O}_i$	Observation space of agent i
R	Total number of federated communication rounds
K	Number of local training steps per round
Policy and Value Parameters	
$\theta_i,\ \nu_i$	Local actor and critic network parameters of agent i
$\theta^{(r)},\ \nu^{(r)}$	Global actor and critic parameters at federated round r
$\theta_i^{\text{fused}},\ \nu_i^{\text{fused}}$	Personalized model parameters after global–local fusion
η	Personalization coefficient balancing local and global policy fusion ($\eta \in [0,1]$)
Reinforcement Learning Terms	
$J_i(\theta_i)$	Expected discounted return (objective) for agent i
$A_i(t)$	Advantage estimate at time t for agent i
$V_i(s_t)$	Value function at state s_t under critic ν_i
$\pi_i(a_t \mid o_t; \theta_i)$	Stochastic policy of agent i parameterized by θ_i
γ	Discount factor controlling reward horizon ($0 < \gamma < 1$)
c_{fb}	Per-parameter computational cost of a forward/backward pass
Federated Learning and Communication Parameters	
n_i	Local data sample size or participation weight of agent i
m	Number of active agents per communication round ($m \leq \lvert\mathcal{N}\rvert$)
b	Bit precision used to transmit model parameters (e.g., 32-bit)
T_{local}	Local compute cost per agent per round
T_{agg}	Aggregation time at the server per round
Differential Privacy Parameters	
C	Gradient clipping norm used to bound sensitivity
σ	Standard deviation of Gaussian noise added for DP
$\varepsilon,\ \delta$	Differential privacy parameters controlling privacy budget
ε_{tot}	Total composed privacy budget over R rounds
q	Client sampling ratio ($q = m/\lvert\mathcal{N}\rvert$)
System and Performance Metrics	
J	Average tri-objective reward combining cost, latency, and security
Cost	Mean operational cost per federated round ($)
Latency	Mean task execution time per round (ms)
Security	Aggregate security score (higher values denote better privacy/compliance)

(Continued)

Table A1 (continued)

Symbol	Meaning
Auxiliary Symbols	
r_t	Immediate reward received at step t
s_t, a_t, o_t	State, action, and observation at step t
$\mathcal{N}(0, \sigma^2 I)$	Gaussian noise distribution with variance σ^2
λ_s	Weight assigned to security in the tri-objective reward function
$Q(s, a)$	Q-function estimating expected reward of action a in state s
Abbreviations and Acronyms	
DT	Digital Twin
MDT	Medical Digital Twin
F-MARL	Federated Multi-Agent Reinforcement Learning
PA3C-PP	Privacy-aware Advantage Actor–Critic with Personalization and Privacy Protection
MA-POMDP	Multi-Agent Partially Observable Markov Decision Process
DP	Differential Privacy
A2C	Advantage Actor–Critic
ICU	Intensive Care Unit
HIPAA	Health Insurance Portability and Accountability Act
GDPR	General Data Protection Regulation
RL	Reinforcement Learning
FL	Federated Learning

References

1. Jameil AK, Al-Raweshidy H. A digital twin framework for real-time healthcare monitoring: leveraging AI and secure systems for enhanced patient outcomes. Disc Inter Things. 2025;5(1):37.
2. Nadeem M, Kostic S, Dornhöfer M, Weber C, Fathi M. A comprehensive review of digital twin in healthcare in the scope of simulative health-monitoring. Digit Health. 2025;11:20552076241304078. doi:10.1177/20552076241304078.
3. Katsoulakis E, Wang Q, Wu H, Shahriyari L, Fletcher R, Liu J, et al. Digital twins for health: a scoping review. npj Digit Med. 2024;22; 7(1):77. doi:10.1038/s41746-024-01073-0.
4. Ooka T. The era of preemptive medicine: developing medical digital twins through omics, IoT, and AI integration. JMA J. 2025;8(1):1–10. doi:10.31662/jmaj.2024-0213.
5. Sun T, He X, Li Z. Digital twin in healthcare: recent updates and challenges. Digit Health. 2023;9: 20552076221149651.
6. Sangaraju VV. AI and data privacy in healthcare: compliance with HIPAA, GDPR, and emerging regulations. Int J Emerg Tren Comput Sci Inform Technol. 2025. p. 67–74. ICCSAIML'25.
7. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. J Health Inform Res. 2021;5(1):1–19. doi:10.1007/s41666-020-00082-4.
8. Ganadily NA, Xia HJ. Privacy preserving machine learning for electronic health records using federated learning and differential privacy. arXiv:2406.15962. 2024.
9. Antunes RS, André da Costa C, Küderle A, Yari IA, Eskofier B. Federated learning for healthcare: systematic review and architecture proposal. ACM Trans Intell Syst Technol. 2022;13(4):1–23.
10. Akhmetov A, Latif Z, Tyler B, Yazici A. Enhancing healthcare data privacy and interoperability with federated learning. PeerJ Comput Sci. 2025;11:e2870.

11. Dang TK, Lan X, Weng J, Feng M. Federated learning for electronic health records. ACM Trans Intell Syst Technol. 2022;13(5):1–17. doi:10.1145/3514500.
12. Pati S, Kumar S, Varma A, Edwards B, Lu C, Qu L, et al. Privacy preservation for federated learning in health care. Patterns. 2024;5(7):100974.
13. Vashisht S, Rani S. Federated learning-driven consumer-centric digital twins for secure and intelligent healthcare 5.0 systems. IEEE Trans Cons Elect. 2026;72(2):4153–62. doi:10.1109/tce.2026.3674985.
14. Feng X, Wu J, Pan Q, Li J. Digital twin enhanced data protection based on Cloud-Edge collaboration in healthcare system. In: 2024 IEEE 9th International Conference on Smart Cloud (SmartCloud). Piscataway, NJ, USA: IEEE; 2024. p. 1–6.
15. Zhang J, Hua Y, Wang H, Song T, Xue Z, Ma R, et al. Fedala: adaptive local aggregation for personalized federated learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press; 2023. p. 11237–44.
16. Tan AZ, Yu H, Cui L, Yang Q. Towards personalized federated learning. IEEE Trans Neural Netw Learn Syst. 2022;34(12):9587–603.
17. Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, et al. Federated learning with differential privacy: algorithms and performance analysis. IEEE Trans Inform Foren Secur. 2020;15:3454–69.
18. El Ouadrhiri A, Abdelhadi A. Differential privacy for deep and federated learning: a survey. IEEE Access. 2022;10(2):22359–80. doi:10.1109/access.2022.3151670.
19. Han J, Wang L, Liu Z, Qin B, Zhang K, Li W. PPFL: privacy-preserving federated learning based on differential privacy and personalized data transformation. IEEE Inter Things J. 2025;12(20):42652–65.
20. Gao J, Li Y. Federated meta-learning for personalized medication in distributed healthcare systems. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Piscataway, NJ, USA: IEEE; 2024. p. 6384–91.
21. Arivazhagan MG, Aggarwal V, Singh AK, Choudhary S. Federated learning with personalization layers. arXiv:1912.00818. 2019.
22. Deng Y, Kamani MM, Mahdavi M. Adaptive personalized federated learning. arXiv:2003.13461. 2020.
23. Dinh CT, Tran NH, Nguyen TD. Personalized federated learning with moreau envelopes. Adv Neural Inform Process Syst. 2020;33:21394–405.
24. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, et al. Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York, NY, USA: ACM; 2016. p. 308–18.
25. Sharafutdinov K, Fritsch SJ, Iravani M, Ghalati PF, Saffaran S, Bates DG, et al. Computational simulation of virtual patients reduces dataset bias and improves machine learning-based detection of ARDS from noisy heterogeneous ICU datasets. IEEE Open J Eng Med Biol. 2023;5(2):611–20. doi:10.1109/ojemb.2023.3243190.